Reengineering the SIPP: The New Dynamics of Economic Well-being System (DEWS)
U.S. Census Bureau, Suitland Maryland
September 27, 2006

This report is released to inform interested parties of ongoing research and to encourage discussion of the dynamics of economic well-being.

The views expressed on statistical, methodological, technical or operational issues are those of the author, not necessarily those of the U.S. Census Bureau**.**

INTRODUCTION

The comprehensive information about individual and household income and program participation collected by the Survey of Income and Program Participation (SIPP) is used by federal agencies to evaluate programs and assess need. A major use of the SIPP has been to evaluate the effectiveness of government programs and to analyze the impacts of options for modifying them. The Social Security Administration (SSA), for example, relies on SIPP data to project baby boomers' retirement incomes and the likely timing of their retirement. The Department of Health and Human Services (HHS) uses SIPP data to evaluate the impact of Welfare Reform and to measure the economic effect of disabling conditions on children and adults.

The SIPP's longitudinal design has many advantages, but imposes considerable burden on respondents and makes review and data processing difficult and time consuming. The re-engineered system, to be known as the dynamics of economic well-being system (DEWS), is expected to reduce respondent burden and attrition and deliver data on a timely basis. Although it will not supply the same level of detail as the SIPP, its design must offer policymakers and researchers data that address the same basic issues. Several options are now being considered for a new system that will provide information on measuring the dynamics of economic well-being.

The DEWS will take advantage of the advances that the Census Bureau has made in acquiring and integrating administrative records with survey data, in modeling local area estimates, and in developing new data collection systems. The new system will eventually supplement data collected from the new survey instrument with administrative records from a variety of sources. The new system can also be used to identify a population cohort that can be measured longitudinally or subsequently surveyed for selected special topics.

SYSTEM GOALS, OBJECTIVES, AND SCOPE

The overall goals of the DEWS is to reengineer the current SIPP to construct a streamlined system that can provide similar information at a reduced cost, with improved data quality, improved timeliness, and improved data accessibility. The system will be able to generate data that can be used, in part, as SIPP data have been used, that is to provide accurate and comprehensive information about the income and program participation of individuals and households in the United States. The DEWS will provide a nationally representative sample that can be used to evaluate the annual and sub-annual dynamics of income, the movements into and out of government transfer programs, and the effect on family and social context of individuals and households. The DEWS will provide this information in a timely manner and at reduced cost through re-engineered survey design, improvements in processing efficiency, and a focused content scope.

The scope of this plan includes those activities that must be undertaken in order to develop the DEWS and to field a new survey instrument in 2009. The first requirement is to identify core elements to be included in the new data system. In addition to identifying core data elements to be included, several configurations of the data collection system are to be examined.

COMMUNICATIONS PLAN

Policy makers, including several major federal agencies, need detailed information on the dynamics of income, poverty, wealth, health insurance coverage, program participation and other aspects of economic well-being. SIPP has provided this information for the past two decades. No other statistical program provides that level of detail. The Census Bureau's plans for a re-engineered system to measure economic well-being will continue to meet the highest priority needs of its historical users, in a timelier manner at reduced cost. To ensure the involvement of SIPP stakeholders, they will continue to be consulted on decisions throughout the development process.

Census management will continuously meet with key stakeholders to assess their priorities for the new system. The Census Bureau has presented overviews of the new system to various audiences, including Council of Professional Associations on Federal Statistics (COPAFS), Census Advisory Committee of Professional Associations, the Congressional Budget Office, Congressional committees, Congressional hearings, and key federal stakeholders. Census staff have organized regular meetings with federal staff and key outside researchers to assess their needs and apprise them of the progress of the new system and will continue this process during the evolution of the system.

SYSTEM OBJECTIVES

The *Systems Development Team* was established to oversee the DEWS and to define the major objectives that were integral to the design, development, and implementation of the new system. The goals and associated objectives are as follows:

- Cost Reduction
  - optimal content design,
  - optimal data collection, and
  - optimal processing system.

- Improved Data Quality – Accuracy
  - test results and evaluation of test results incorporated into design and project plan,
  - lower attrition rates,
  - integrated quality assurance, and
  - reduce seam bias across panels.

- Improved Data Quality – Timeliness
  - finalized core design components by 2007,
  - release data faster than SIPP, and
  - information products and services disseminated in alignment with customer and stakeholder needs.

- Improved Data Quality – Accessibility
  - publicly available microdata files,
  - information products and services disseminated in alignment with customer and stakeholder needs, and
  - strategic partnership with stakeholders.

An inter-divisional project group was established to oversee each of these objectives. Below is the progress to date for each group.

I. ADMINISTRATIVE RECORDS

A. Administrative Records and Survey Data Prototype Development

The *Prototype Development Group* is responsible for all phases of the design and development of integrated current and proposed survey and administrative records-based prototype systems and resulting data sets. Prototype development is a continuing process that will evolve along with the entire system. The early prototypes will be used mainly to determine disclosure issues associated with creating a Public Use File containing administrative records data. Of particular importance in this activity is the evaluation of the use of administrative records as a measure of data quality and for use in edits and imputations. At every phase of development, the Census Bureau's Disclosure Review Board will review the resulting data files.

The initial survey records/administrative data prototype was a combination of one year of data from the March 2001 Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS) and Medicaid data from the 2001 MSIS. The second prototype

consisted of two years of ASEC data and Medicare data[1] from the MEDB extract file. The third prototype, currently under development, will be 2004 wave 1 SIPP matched to demographic, public housing assistance, and industry codes administrative data.

B. Administrative Records to SIPP Assessment

The *Administrative Records to SIPP Assessment Group* is involved in ongoing evaluations to determine the current availability of administrative data sources that have sufficient national coverage and can be obtained in a consistent and timely manner for inclusion in the new survey processing. These evaluations are iterative, and will continue as the design progresses and the new survey content becomes more solidified. As a first step the group attempted to align core variables from 1993 SIPP Panel longitudinal file with currently available administrative records data including identifying the variable(s), defining the coverage of the data source, and determining the lag between when the administrative data are collected and when they are available at the Census Bureau. Using this 1993 file as a starting point is advantageous because it limits the scope and the number of items to those previously used for longitudinal data analysis. Typical match rates for administrative data to survey data are in the low 90 percent range. Examples, by core section, are discussed below.

Health Insurance

- *Medicare* — The majority of SIPP items align with a current administrative records source and a national-level file is available in April of each calendar year.

- *Medicaid* — SIPP Medicaid items align with a current national-level administrative records source. However, there is a 3-year lag in obtaining this data. There are no administrative data sources for Medicaid state expansion or other types of public health insurance.

Assets

- *Asset ownership, gross income, and net income* — All aggregate SIPP asset items align with current Internal Revenue Service (IRS) administrative records sources—IRS 1040, 1099-INT, and 1099-DIV. These national-level files have negligible delays in delivery to Census. However, obtaining access may be a time and resource intensive activity

Labor Force

- *Total earnings* — Total earnings from a job(s) or income (loss) from a business(es) align with administrative data available from

[1] Medicaid data was dropped from the second prototype because the time lag for availability of the MSIS is too long.

	the IRS 1040. Access to these data have the same constraints as for assets.
Demographics	
• *Demographic characteristics*	Data from the Census Numident and the Person Characteristics File (PCS), both national files, are available for some respondent characteristics (e.g., age, Hispanic origin, race, sex, U.S. citizenship).

Many programs are administered at the state level. At the current time, our access to state data to address receipt of these programs is limited. However, this is an area that we will look at closely to determine which state partnerships can be developed in order to obtain the necessary data files in the future.

C. Administrative Records and Public Use Data Research

As previously indicated, identifying and mitigating any disclosure issues associated with creating microdata based on a combination of administrative records and survey data is an important area of ongoing research. The data user community indicated that it is critical that they have access to public use microdata from the DEWS without having to travel to a Census Bureau Research Data Center. Therefore, the Census Bureau is working with the National Academy of Sciences to have the Committee on National Statistics (CNSTAT) convene a panel of experts in the fields of survey methods research, administrative records, data linkage techniques, income measurement, social welfare program policy analysis, and confidentiality protection. The charge of this expert panel will be to provide an evaluation, as well as their final recommendations, (in the form of a published report) addressing: (1) the costs/benefits of various strategies for data linkage, (2) the accessibility of relevant administrative records, (3) the operational feasibility of linking administrative records and survey data, (4) the quality and usefulness of the linked data, and (5) strategies for providing public access to the linked data while protecting the confidentiality of individual respondents as it relates to the DEWS.

Additionally, the Census Bureau's administrative records staff are currently collaborating with other federal agencies on several projects involving SIPP data, administrative records, and public use data. A brief synopsis of these initiatives follows:

1. Understanding and Increasing the Utility of SIPP/SSA Data through the Congressional Budget Office (CBO) Long-Term Retirement, Health and Disability Modeling as well as Developing a PUF based on a combination of SIPP and SSA Data.

   The Census Bureau, SSA, and CBO are collaborating in the development of a Public Use File (PUF) containing SIPP data linked to SSA administrative data (including SSA files that contain Federal tax information). A long-term modeling effort that

began in 1999 has already generated several new tools for studying policy issues such as Social Security and Medicare. This project extends this work by estimating individual-level models of retirement decisions, marital transition, patterns of husband-wife matching across demographic and economic characteristics, labor force participation; developing methods for projecting individual earnings for future cohorts of workers and retirees; as well as studying the relationship between health, disability, and earnings.

The project involves 4 parts:

A. Build the Gold Standard - this involves extracting variables from all the panels, standardizing them, and then merging on the administrative data. From SSA, we receive and merge Average Indexed Monthly Earnings (AIME - a summary measure of lifetime earnings used to calculate SSA retirement and disability benefits), the PIA (Primary Insurance Amount - derived from AIME), the MBA (monthly benefit amount), and the TOB (type of benefit - reason for getting benefits). Jointly from IRS and SSA , we receive and merge total earnings between 1937 and 1951, and annual earnings from 1951 to 2003. From 1951 to 1977, annual earnings are capped at the FICA max. Beginning in 1978, total earnings are uncapped. These earnings data come from the W-2 forms filed by employers and NOT from the IRS Form 1040.

B. Synthesize SIPP and IRS/SSA data. An extensive set of programs that synthesize all approximately 500 variables from the Gold Standard file were created.

C. Test analytic validity. After synthesizing, running regressions and calculating other statistics of interest, comparisons are made between the synthetic and original data. If the two sets of data are not comparable, re-evaluation and re-synthesizing are required.

D. Test for disclosure risk. Sam Hawala of SRD and a John Abowd graduate student at the Cornell RDC have attempted to re-identify respondents in the public use SIPP files, so far unsuccessfully. IRS, SSA, and Census DRB must agree to a public use file release.

2. <u>SIPP–School Enrollment and Financing</u>

The goal of this project is to improve the quality of the School Enrollment and Financing topical module data by linking to a PUF of the Integrated Post-Secondary Education Data system (IPEDS), a survey conducted by the National Center for Education Statistics (NCES). This SIPP topical module records were matched with IPEDS and characteristics of the SIPP respondent's post secondary school (size, tuition, percent minority, etc.) were obtained and added to the topical module file.

3. SIPP for SSA Beneficiaries Study

This project administered the Wave 6 instrument from the 2001 SIPP panel to a separate sample of current Supplemental Security Income (SSI) and Social Security Disability Insurance (SSDI) beneficiaries. The instrument contains the core items, and the topical module items on assets, liabilities, and eligibility; medical expenses and utilization of health care by adults and children; work-related expenses; and child support paid. The sample is selected randomly by ZIP code by the SSA using sampling guidance provided by the Census Bureau, and is designed to yield approximately 1000 interviews for each SSA program. This project supplies a micro-data file to SSA, supplementing existing SIPP beneficiary lists.

The collective results of the CNSTAT panel and these initiatives will form the basis for a DEWS data dissemination strategy that will meet the needs of the data user community while ensuring the confidentiality of our respondents.

II. SURVEY CONTENT, METHODOLOGY, AND DESIGN

A. Survey Content

The *Content Group* will determine the variables collected in the new DEWS survey instrument. The group is using the 1993 SIPP Panel longitudinal research file as the starting point. Additional variables will be added and/or current variables eliminated based on the outcome of stakeholder meetings and Census analytical requirements. Another source of information on core data usage was a survey conducted by the Census Bureau which indicated that 91 percent of users access demographic and household composition variables, 73 percent access labor force and transfer program variables, 65 percent access job and business earnings, and 60 percent access other sources of income. Other factors, such as budget constraints, and the final survey methodology—including mode of data collection (which is discussed in the next section), will also impact the final survey content.

Because there have been changes in SIPP core content since this 1993 longitudinal file, a brief synopsis of each section of the 1993 longitudinal file follows:

Labor force participation and income (e.g., current employment status, weeks worked, weeks spent looking for work or on layoff); if ***employed by a job or jobs***—the number and name of each employer, type of business or industry, the type of work, the frequency and amount of pay. If ***owned a business or businesses***—the type of business and work, the number of hours worked per week, the total number of employees; r***eceipt and***

amounts from general income sources (e.g., retirement, pension, disability, Social Security, Medicare, Supplemental Security Income (SSI)); ***public and private health insurance coverage***; ***asset ownership and income***; and, ***school enrollment and assistance***.

A series of matrices have been developed to help organize the content discussion. Input was solicited from three areas in order to derive the final survey content—they are: key stakeholders, Census data analysts, and administrative records staff. The goal of these matrices is to provide a complete picture of the user's needs and current availability of data. Of particular importance in this information compilation is an assessment of which data items need collection sub-annually and which do not. As will be noted later, many sub-annual data items will be collected within the Event History Calendar (EHC).

Because of budget restrictions and concerns about respondent burden, our vision is that the new survey will also serve as a screening device for future topic surveys. In other words, two or three questions about child well-being in the new survey could identify the universe of respondents for a follow-on survey on the well being of children. Once established, the core content of the new survey should remain constant.

B. Methodology and Survey Design

In conjunction with the Content Group, the *New Survey Group* is leading the development of the sample frame, structure, and components for a new data collection instrument. The group is developing potential options under the following assumptions:

- The system will collect data covering at least three years and have the ability to provide sub-annual data for that period.
- Timeliness of the data file release is critical.
- Sampling, estimation, and disclosure-proofing strategies can be developed to support any options being considered.

1. **Level of Detail**

The historical precedent set by the SIPP program is that of detailed data on a core set of questions, with topical modules of a broad and varied scope added for each wave of interviewing. This scope and detail contributed to the complexity of the instrument, the length of the interview, and subsequent high attrition rates, all of which we hope to reduce with the new survey. The goal of reducing the content detail will be to produce a more streamlined survey that will be less costly and reduce respondent burden.

2. **Periodicity and Question Format**

The current SIPP program administers three waves or interviews in each year. The interview rotation design means that data are collected in each of the 12 months of the calendar year in order to level field resource requirements. This requires a constant level of production-readiness which encumbered the preparatory and the post-collection processes, contributing to data products that were not timely and were also error-prone.

After consideration of the existing 4 month, a six month, or an annual recall period and the costs associated with each, it was decided to consider a single annual survey contact in the household with a reference period long enough to capture retrospective information on sub-annual dynamics. However, this requires a non-traditional approach to obtain these month-to-month transitions—such as an EHC. Research has shown that EHCs improve the accuracy of life event reporting by facilitating a more natural conversational framework and by emphasizing the inter-relationships of events.[2]

To accomplish this, discussions are ongoing with the University of Michigan, who developed the EHC for the Panel Survey of Income Dynamics (PSID). While we believe this approach ultimately holds the most promise for the DEWS data collection instrument because of the improved accuracy of the transition estimates as well as cost savings (in terms of interviewing costs), this methodology deviates from current Census Bureau methodology—in terms of both instrument development[3] and interviewing procedures. Therefore, it is imperative that we pursue this option carefully.

In addition to PSID, several other surveys are using the EHC approach. These include:

- National Survey of Family Growth – CDC/NCHS
- National Survey of Adolescent Health (AddHealth) – University of North Carolina
- Los Angeles Family and Neighborhood Survey – UCLA
- English Longitudinal Study of Ageing – NatCen UK, University College London, and the Institute for Fiscal Studies

---

[2] Wiebe, E. and Landis, K. Evaluation of an Electronic Event History Calendar, RTI.

[3] All data collection instruments must be in a format that will successfully interact with our case management systems. The EHC developed for the PSID is a visual basic interface with an ACCESS database. The interaction of this type of instrument is currently being tested at the Census Bureau to determine the implications to either case management or ROSCO.

- Health and Retirement Study – University of Michigan (National Institute on Ageing)
- 1998 National Retrospective Demographic Survey (EDER) – Mexican Census Bureau (INEGI)

To begin design of a survey instrument that incorporates an EHC to obtain sub-annual estimates of events with typically short term spells, the potential content have been assigned to various concepts—those that do no require sub-annual reporting to create meaningful estimates (e.g., assets) and those that do (e.g., receipt of and amount of means-tested welfare programs).

3. **Mode**

Traditionally, SIPP has relied on personal visits by field staff to establish and maintain rapport with the sample cohort, although over the life of a longitudinal panel, telephone interviews are encouraged in order to minimize costs. The mode for the new data collection system is highly dependent on the complexity of the questionnaire, the budget, and the desired response rate.

4. **Sample Source and Size**

There are several options for generating a sample frame[4], each with different characteristics and advantages. These options include:

- sampling from the Master Address File (MAF) with auxiliary data from another source (ex: ACS or administrative records);
- sampling directly from ACS interviewed cases;
- using the existing SIPP sample[5] already identified and unduplicated.

The sample size will be determined by the level of reliability required for specific estimates as well as the survey's budget. The sample size required to produce only reliable national estimates is much smaller than what would be needed to produce reliable state estimates even for only the largest states.

[4] These options will be discussed more fully in the sample design section of this document.

[5] The 2000 sample redesign selected enough SIPP cases to cover the 10-year period between decennial censuses. The SIPP sample cases that have not been sent to the field are available for the dynamics of economic well-being system.

5. **Following Movers**

All of the options being discussed involve following movers. The methodology will be similar to what is currently being done in the SIPP.

6. **Survey Design**

The current SIPP design includes an oversample for low-income households. This may still be a requirement for the DEWS, but could shift to oversampling of other subgroups (the elderly, families with children, or race/Hispanic subgroups). For any oversampled group, auxiliary data are needed to determine which cases are in the oversample group. If oversampling is based on a characteristic that changes over time, greater sampling efficiencies will be possible by interviewing within the shortest time possible following the sample selection. For example, sampling directly from ACS will likely result in the highest oversampling efficiency, while sampling from the MAF will be a little less efficient.

The current SIPP design is clustered into Primary Sampling Units (PSUs) for efficiencies in field representative workloads. In order to continue with this PSU type design, the sampling frame will need to contain a sufficient number of cases within those sampled areas in order to ensure equitably sized workloads or, alternatively, the size of the PSUs will need to be adjusted.

Finally, the coverage of the frame will need to be taken into account in evaluating the alternatives sources for the sample.

C. Post Data-Collection Processing

The *Data Processing Team* has been formed and is charged with developing a new SAS-based system to process the DEWS data and create and disseminate the internal and public use data products. Information concerning team membership and specific tasks are being developed.

III. DATA PRODUCTS

The primary product of the DEWS, as currently planned, is a series of micro-data files that will be publicly released in a timely manner. Our original intention was to create a two-year retrospective file based on an existing demographic survey and enhanced with current administrative data to be released in December 2008. However, comments received at the June 8, 2006 stakeholder meeting at the Brookings Institute indicated that this file would be of little value to users. Therefore, we are re-evaluating the initial data product from the DEWS.

The *Data Products Group* will work in parallel with the Survey Content, Survey, and Data Processing groups to determine the actual data products produced from the new system. This determination will be based primarily on the needs of our data users, both internal and external and any disclosure issues identified by the Census Bureau's Disclosure Review Board. While the Data Products Group has been established, work in earnest will not commence until the survey methodology is more clearly defined. It is, however, currently expected that data products will include all of the following: typical public use microdata files, restricted internal files available at the Census Bureau and RDCs, and public use files containing synthetic data components.

Money was allocated in the 2007 House mark for continued data collection for 2004 Panel during FY 2007. Therefore, Waves 9 and 10, and Wave 11 if funding permits, core will be collected. As the table in Attachment A shows, the last data set of deliverables from the 2004 SIPP Panel will be released in 2007 and early 2008 (December, January, and February for the Wave 8 topical modules). In addition, the Census Bureau will continue to produce analytical reports evaluating the 2004 SIPP Panel. Current SIPP P70 reports can be found on the Census website at http://www.sipp.census.gov/sipp/p70/p70s.html. This collection includes a series of report on the Dynamics of Economic Well-being (e.g., P70-105, P70-100).

IV. **STAKEHOLDER SUPPORT**

Stakeholder support is crucial to the success of this new system. We envision this as a collaborative effort between the Census Bureau and its many stakeholders during the lifetime of the project. Meetings with each individual agency to determine its needs are ongoing. As the development of the system progresses, we will continue to have inter-agency meetings to discuss our progress as well as any content issues for inclusion or deletion from the survey that have arisen.

CONCLUSION

During the life of the development of and eventual maintenance of this project, the structure of the interview, whether it requires personal visit interviewing, the number of re-interviews, whether movers are followed, and the requirement to measure program participation and eligibility must be balanced with cost and sample size considerations. In addition, the timeliness and utility of data production and release with respect to collection will be areas needing continuous attention and review. Among the objectives for this re-engineering are improvements in the timeliness of data releases with respect to data collection, and improvement of processing efficiency by simplifying the structure of the files and the data collection.

Attachment A

2004 SIPP Panel Data Products Release Dates

Wave	Core Release Date	Topical Module ™	TM Release Date
1	10/2006*	Recipiency History	10/2006
		Employment History	10/2006
2	11/2006	Work Disability History	11/2006
		Marital History	11/2006
		Fertility History	11/2006
		Household Relationships	11/2006
		Migration History	11/2006
		Education and Training History	11/2006
3	1/2007	Medical Expenses/Utilization of Health Care	11/2006
		Work-related Expenses	11/2006
		Child Support Paid	11/2006
		Child Well-being	9/2006
		Assets, Liabilities, and Eligibility	1/2007
4	3/2007	Annual Income and Retirement Accounts	12/2006
		Taxes	12/2006
		Work Schedule	12/2006
		Child Care	12/2006
5	4/2007	Adult Well-being	3/2007
		School Enrollment and financing	4/2007
		Child Support Agreement	4/2007
		Support for Non-household Members	4/2007
		Functional Limitations and Disability – Adult	4/2007
		Functional Limitations and Disability – Child	4/2007
		Employer Provided Health Benefits	4/2007
6	5/2007	Assets, Liabilities, and Eligibility	5/2007
		Medical Expenses/Utilization of Health Care	5/2007
		Work-related Expenses	5/2007
		Child Support Paid	5/2007
7	6/2007	Annual Income and Retirement Accounts	8/2007
		Taxes	8/2007
		Retirement and Pension Plan Coverage	10/2007
		Informal Care-giving	10/2007
8	7/2007	Welfare Reform	12/2007
		Child Well-being	1/2008
		Child Care	2/2008

* Re-release of Wave 1